

## CS378 Fall 2022 Midterm Topics

The midterm is a **closed book in-class exam**. You will have 75 minutes to complete it.

The exam will have the same broad types of questions (multiple choice, short answer, long answer) as past midterms, with a greater emphasis on open-ended and conceptual questions rather than algorithm execution. Do show your work, but for our sake and yours, please try to refrain from writing an essay as a response unless the question explicitly asks for it.

**Topics** Below is the list of topics that will be covered on the midterm, which is the bulk of the course material thus far.

- Bag-of-words features: how these feature spaces look and how they work for classification
- Perceptron (binary): algorithm, loss function
- Logistic regression (binary): model (the LR formula), training objective, gradient update
- Sentiment analysis: what kinds of features are useful?
- Multiclass classification: how weights and features work in this setting
- Multiclass perceptron: model definition, how to train it
- Optimization: stochastic gradient descent, impact of step size on optimization, impact of initialization (particularly for FFNNs)
- Feedforward neural networks: definition, initialization
- Training neural networks
- Word embeddings: skip-gram model, definition, properties, potential for bias
- Deep averaging networks: model from Assignment 2, limitations of the model
- POS tagging: understanding ambiguities (like the *Fed raises...* example)
- Sequence labeling as classification: how to build position-sensitive features
- Hidden Markov Models: definition, parameter estimation (counting and normalizing), Viterbi algorithm
- Beam search
- Constituency syntax: what trees look like, ambiguities (*Ban on nude dancing..., eat spaghetti with chopsticks*)
- PCFGs: definition, parameter estimation (counting and normalizing), what binarization is (you won't be expected to do it), CKY algorithm
- Dependencies: definition, differences from constituency. Shift-reduce won't be a major focus as it's covered so close to the midterm

**Other content** You should expect to see examples of text and be comfortable reasoning about how these algorithms might work on such examples, as in the assignments so far. We won't expect you to know things like part-of-speech definitions or have encyclopedic knowledge of grammar structures—we will provide the necessary information for such questions.

**Readings** We won't expect you to know content from the Eisenstein book or other readings that hasn't been covered in lecture or assignments.

**Practice problems** The best source of practice problems is the last couple of midterms, which are posted on the course website. Note that Naive Bayes classification (spring 2019 multiple choice Q1-4) has been dropped as a topic and the focus may have changed from prior courses, so if a topic seems unfamiliar, it may not have been discussed this time around.